2024-10-22

October 21, 2024

1 Neural Networks

Prof. Forrest Davis

1.1 Mathematical Basics

1.1.1 Dot Product

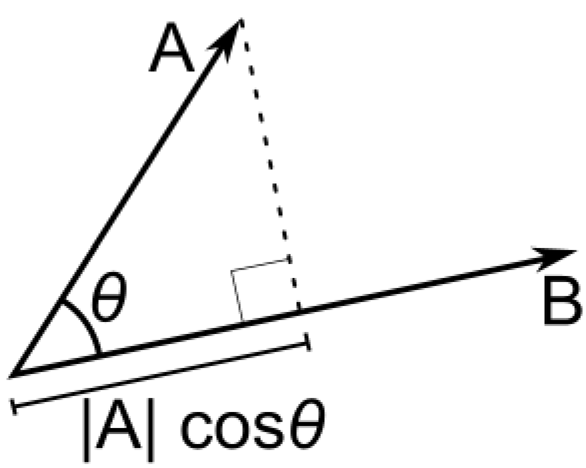
The dot product is defined between vectors of equal length n as:

$$\mathbf{v} \cdot \mathbf{w} = \sum_{i=0}^{n} \mathbf{v}_i \mathbf{w}_i = \mathbf{v}_0 \mathbf{w}_0 + \mathbf{v}_1 \mathbf{w}_1 + ... + \mathbf{v}_n \mathbf{w}_n$$

That is, the dot product of two vectors yields a scalar (for a basic refresher on dot products, see this video). We can interpret this scalar as a measure of the angle between the vectors. Formally, the dot product, in geometric terms, is:

$$\mathbf{v} \cdot \mathbf{w} = |\mathbf{v}||\mathbf{w}|\cos(\theta)$$

Where $|\mathbf{v}|$ is the magnitude (length) of the vector $\mathbf{v}$. Recalling the properties of the cosine function, it is easy to see that the dot product between two perpendicular vectors is 0, and the dot product between two vectors which face the same direction is $|\mathbf{v}||\mathbf{w}|$. In other words, the dot product grows as the vectors are more aligned in space.



1.1.2 Matrix Multiplication

Matrix multiplication builds on the dot product as a means of multiplying matrices together (see this video series for a refersher on matrix multiplication).

Formally, suppose we have two matrices:

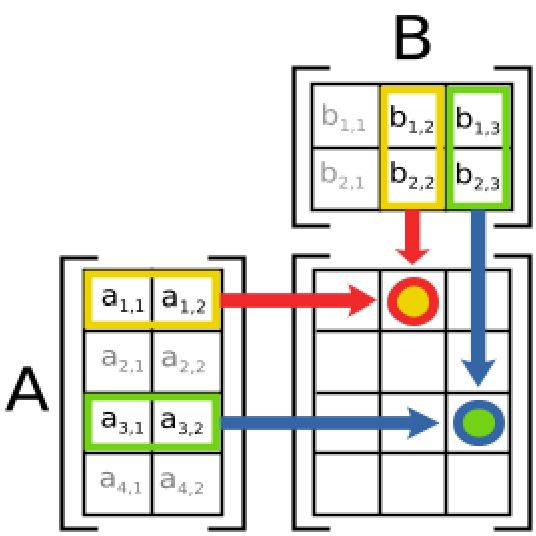
$$\mathbf{A} = \begin{pmatrix} a_{(0,0)} & a_{(0,1)} & a_{(0,2)} \\ a_{(1,0)} & a_{(1,1)} & a_{(1,2)} \\ a_{(2,0)} & a_{(2,1)} & a_{(2,2)} \end{pmatrix} \quad \text{and} \quad \mathbf{B} = \begin{pmatrix} b_{(0,0)} & b_{(0,1)} & b_{(0,2)} \\ b_{(1,0)} & b_{(1,1)} & b_{(1,2)} \\ b_{(2,0)} & b_{(2,1)} & b_{(2,2)} \end{pmatrix}$$

Matrix multiplication would return a new matrix:

$$\mathbf{C} = \begin{pmatrix} c_{(0,0)} & c_{(0,1)} & c_{(0,2)} \\ c_{(1,0)} & c_{(1,1)} & c_{(1,2)} \\ c_{(2,0)} & c_{(2,1)} & c_{(2,2)} \end{pmatrix}$$

where:

$$\begin{aligned} c_{(i,j)} &= \sum_{k=0}^{2} a_{(i,k)} b_{(k,j)} \\ &= a_{(i,0)} b_{(0,j)} + a_{(i,1)} b_{(1,j)} + a_{(i,2)} b_{(2,j)} \end{aligned}$$



As you might recall, matrix multiplication has a shape restriction. To multiply two matrices **A** and **B**, the shape of **A** must be n X m and the shape of **B** must be m X k. That is, the inner dimensions must align.

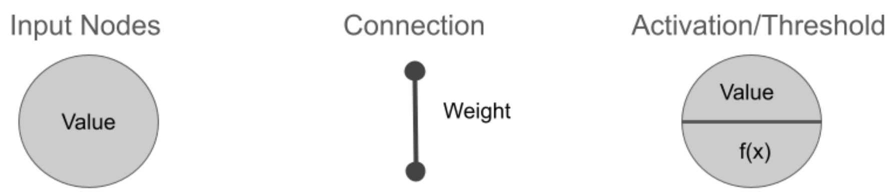
Notice, that matrix multiplication is nothing more than the dot product between the rows in **A** and the columns in **B**. So we can use matrix multiplication as a way of doing many dot products between vectors.

1.1.3 Transpose

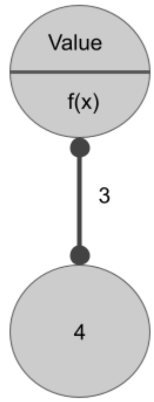
Recall that transposition of a matrix flips the matrix over its diagonal. Courtesy of wikipedia, you can see this here

1.2 Graphical Representation of Neural Networks

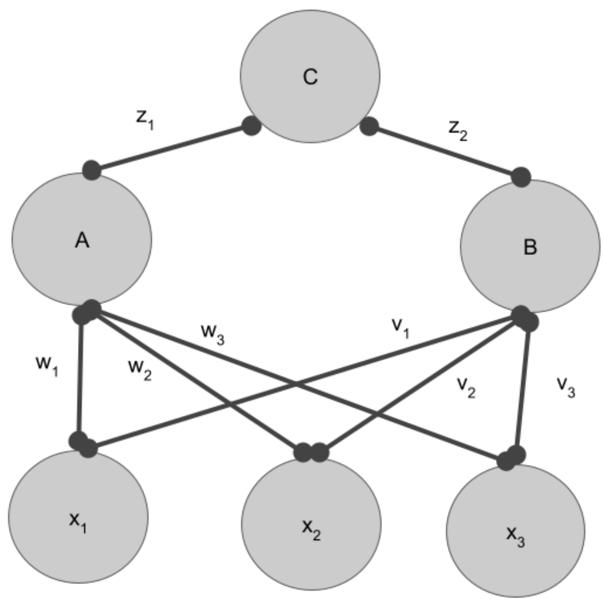
- A neural network is comprised of three objects: nodes (which take values), connections (which link nodes with weights), and activation/threshold functions (which apply to nodes)
- We will represent these three things graphically as below:



- For example, the following graph shows an input node with a value of 4 connected to another node with an activation function applied. The connection has a weight of 3, and the activation function is the identity function (i.e. $y = x$).



1.3 Multi-Layer Perceptrons



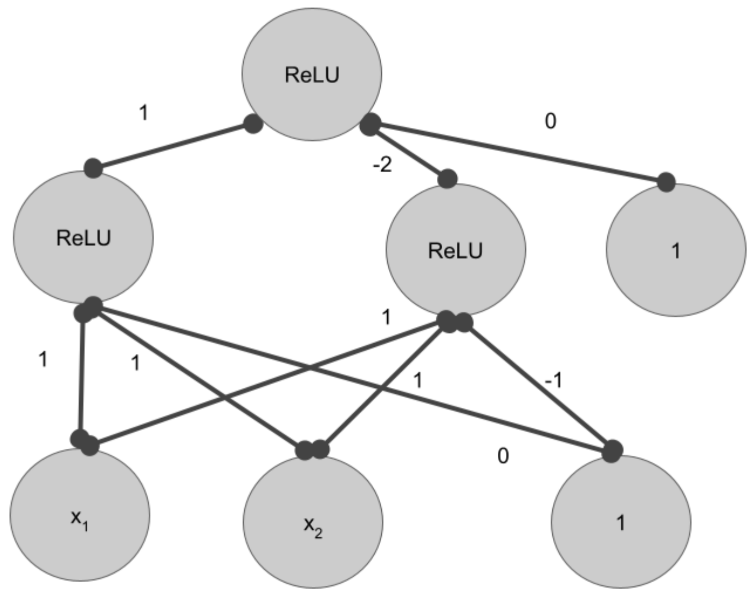
Question - What is the value of C (as an expression)?

- What do you notice?

Question - What does this tell us about how activation functions relate to learning? In other words, how do activation functions relate to the class of problems I can learn? Hint: What was my assumed activation function here?

1.4 Non-Linear Activation Functions

- Consider the following graph with a new activation function called `ReLU`.
- `ReLU` is defined as the element-wise application of $\text{ReLU}(\vec{x}) = \max(0, x_i)$
- For example, $\text{ReLU}([-1, 9, -10, 2]) = [0, 9, 0, 2]$



Question: Translate this graph to an equation with matrices and vectors

- We are interested in a dataset $X^{4\times 2}$ (i.e. a dataset of 4 samples each with 2 features):

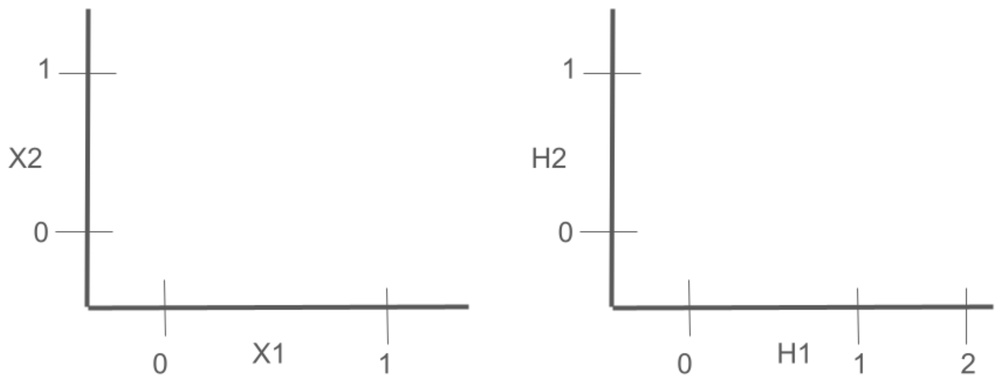
$$\begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}$$

Question: Calculate H_1, H_2, and O for each sample.

Question Fill in this table with your results

x_1	x_2	H_1	H_2	O

Question: Fill in these two plots with information from your table. Labels points with X if their output is 0 and with Δ otherwise.



Question Can you draw a line seperating the classes?

Question: What is the hidden layer doing?